Simulation study on effects of signaling network structure on the developmental increase in complexity

Soile V. E Keränen
Department of Genomic Sciences
Ernest Orlando Lawrence Berkeley National Laboratory
MS 171-84, 1 Cyclotron Road
Berkeley, CA 94720
USA

Tel: + 1 – 510 – 486 5662
Fax:
email: svekeranen@lbl.gov

Corresponding author:

Soile V. E Keränen
Department of Genomic Sciences
LBNL, MS 171-84
1 Cyclotron Road
Berkeley, CA 94720
USA

Tel: + 1 – 510 – 486 5662
Fax:
email: svekeranen@lbl.gov

Key words:
Signaling networks, complexity, network structure, development, cell types

Abstract

The developmental increase in structural complexity in multicellular lifeforms depends on local, often non-periodic differences in gene expression. These depend on a network of gene-gene interactions coded within the organismal genome. To better understand how genomic information generates complex expression patterns, I have modeled the pattern forming behavior of small artificial genomes in virtual blastoderm embryos. I varied several basic properties of these genomic signaling networks, such as the number of genes, the distributions of positive (inductive) and negative (repressive) interactions, and the strengths of gene-gene interactions, and analyzed their effects on developmental pattern formation. The results show how even simple genomes can generate complex non-periodic patterns under suitable conditions. They also show how the frequency of complex patterns depended on the numbers and relative arrangements of positive and negative interactions. For example, negative co-regulation of signaling pathway components increased the likelihood of (complex) patterns relative to differential negative regulation of the pathway components. Interestingly, neither quantitative differences either in strengths of signaling interactions nor multiple response thresholds to signal concentration (as in morphogen gradients) were essential for formation of multiple, spatially unique cell types. Thus, with combinatorial code of gene regulation and hierarchical signaling interactions, it is theoretically possible to organize metazoan embryogenesis with just a small fraction of the metazoan genome. Because even small networks can generate complex patterns when they contain a suitable set of connections, evolution of metazoan complexity may have depended more on selection for favourable configurations of signaling interactions than on the increase in numbers of regulatory genes.

Introduction

The evolution of complex multicellular life requires the ability to generate differentiated cell types in a spatially and temporally co-ordinated manner. Developmental increase in complexity during metazoan morphogenesis is guided by developmental regulatory genes, which are expressed in complex but stereotypical spatial and temporal patterns. The expression patterns depend on regulatory interactions between genes. These interactions depend on the molecular affinities of each gene product to other gene products or to regulatory DNA sequences. All genes are regulated by products of the same and/or other genes. Therefore, the interactions form an information processing network coded within the organismal genome. Though genome function is usually studied in the context of single cells (e.g., Guelzim et al., 2002; Ravazs et al., 2002; Snel et al. 2002) in multicellular animals this information network spans all cells (Gerhard and Kirschner, 1997). Key questions in the evolution of multicellular life are; how does a gene interaction network generate multiple cell types, each expressing different but overlapping parts of the genome, and what features of such network affect the morphogenesis of complex organisms.

The development of complexity has traditionally been simulated through spatio-temporal changes in binary fate choises that arise in a self-organizing manner (e.g., Turing, 1952; Collier et al., 1996; Salazar-Ciudad et al., 2000). These models often emphasize the generation of periodic patterns, which can explain the development of repetitive structures, such as hydra tentacles (Turing, 1952), pigment spots (Asai et al., 1999) or animal segments (Salazar-Ciudad et al, 2001a). However, metazoan bodyplans contain many non-periodic features. Development of unique organs often requires complex, and effectively nonperiodic pattern formation. Even serially repeated organs, like segments, often have unique, non-periodic differences (McGinnis and Kuziora, 1994; Gilbert, 1998). Moreover, a unicellular yeast has ~6000 genes to maintain its single cell, whereas a metazoan fruitfly has only ~14000 genes for differentiating its thousands of embryonic, larval and adult cell types in a complex and stereotypical pattern (Rubin et al., 2000). To answer this N-value paradox,. i.e., why morphological complexity of the organism does not better correlate its gene numbers (Claverie, 2001), we need to know: what kind of genomes are required to generate multiple different cell types, and how much input information is required for multi-cell type pattern formation?

The role of individual gene-gene interactions' strengths, i.e. the molecular kinetics and affinities, in pattern formation has been extensively studied (e.g., Asai et al., 1999; von Dassow and Odell, 2002). Therefore, I asked: what is the role of network structure (i.e., the graph that describes the topology of gene-gene interactions or the qualitative connections between genes) in generating patterns, especially the complex, nonperiodic ones that are essential for metazoan development, and what kinds of features are involved in generating developmental complexity?

The 14000 genes of a fruitfly have 14000^2 potential two gene interactions and 2^{14000} potential binary expression combinations. The real complexity of connections between the genes is even more hyperastronomical, and largely unknown, involving, e.g., multiple levels of post-transcriptional regulation (Gerhart and Kirschner, 1997), multiprotein complexes (e.g., Sobott and Robinson, 2002), various forms of second messenger signaling (Hunter, 2000), and purely

mechanistical integration of morphogenetic effects of various signaling pathways (Gerhart and Kirschner, 1997; Oster and Murray, 1980; Salazar-Ciudad and Jernvall, 2002). This makes *in vivo* biological networks effectively unmodelable. Hence, simpler models must be used for initial mapping of the generic pattern formation properties of information processing networks.

I began the modeling with extremely simple genomes of the virtual organisms containing only 2-, 3-, 4-, 5-, or 6-pairs of genes (an intercellular signal and its specific intracellular effector), and coded for the 4 – 36 potential interactions between each signal and all its targets. The model was based on several premises.

1) Each cell in an organism has the same genome. Since each cell is a separate compartment, local differences in signaling can lead to differentiation of cell types (Collier et al., 1996). Inversely, because different cells are individual compartments, they can have different proteomes (Gerhart and Kirschner, 1997). Thus, each cell can access different parts of the genome when encountering the same signals. Though the coded strengths of the potential interactions remain constant through the simulation, the actual signaling depends on the concentrations of signal and effector gene products in each cell.

2) Each signaling interaction affects directly, additively, and independently to the concentrations of the signals and the effectors in the target cell.

3) Because many important patterning events occur at blastoderm stage, when the main embryo in some metazoans, like fruitflies or sea urchins, is essentially a simple hollow two-dimensional sheet of cells (Gerhart and Kirschner, 1997; Wolpert, 1998), formation of complex patterns does not need to occur in a morphologically complex environment. Consequently, the virtual organisms in the model mimicked a hollow cellular blastoderm with a virtual organism consisting of a closed 2D-sheet of 30 x 40 cells.

4) Though many of the extant models generate nonperiodic patterns by differential responses to different concentrations of a diffusible morphogen from a local source (Wolpert, 1998; Gurdon and Bourillot, 2001), pattern formation can also be based on cell-to-cell relays, which is a common form of intercellular communication (e.g., Bosenberg and Massague, 1993). Hence, in this model, the intercellular signaling cascades spread deterministically in two-dimensions via cell-to-cell relays.

5) The formation of complex spatial patterns requires at least one spatial inhomogeneity that can either arise randomly by stochastic phenomena (Turing, 1952; Elowitz et al., 2002) or be developmentally or environmentally determined (Wolpert, 1998; Bachvarova, 1999). The signaling cascades in the model began from a point source of one signal, that acted as an initial input. As an intercellular signal was not effective unless the target cell expressed its specific effector, all cells expressed the effector to the input signal at the beginning.

Results

Simple multisignal networks with negative interactions can translate a single signal to multiple cell fates

The first question was: how easily do the identical cells with same genome can differentiate into more than one cell type, when presented one spatial input? To answer this, I generated 2-, 4-, 5-, and 6-signal genomes, and tested their pattern formation capabilities in 30 x 40 cell virtual organisms that mimicked metazoan blastoderm embryos. Though the genes were the same for all 1000 networks in each simulation, the networks differed by the quality (positive or negative) and positions of the connections, and the quantitative strengths of each connection. The temporal dynamics of gene expression were iterative. At the beginning, the cells of an organism expressed initial input, which was used for reading the genome into an output of the 1st iteration. During following iterations, the output of the previous iteration became the new input.

In any networks, if all cells expressed initially the same selection of genes, patterns were not formed, though the combination of genes expressed in all cells may have changed from the initial one (not shown). This confirmed the knowledge that at least one spatial difference is needed to initiate the pattern formation cascade (Turing, 1952). Figure 1 shows how one initial input signal in one cell is converted into a complex pattern by relay of secondary signals and effectors, which become induced in different combinations. In other words, developmental increase in complexity can begin from very simple inputs.

How easily does the formation of complex patterns occur? Figure 2A shows how many out of 1000 simple networks produced any expression patterns, and the summed complexities of the formed patterns. The pattern formation frequencies and the pattern complexities increased with the size of the network. Though no pattern or simple patterns (≤2 gene combinations) are most common, complex patterns (>3 different gene combinations) exist at reasonable frequencies. Because the networks above were both simple (≤12 genes, ≤6 signaling inputs to control each gene), the evolution of developmental signaling networks capable of generating complex outcomes was probably not limited by rate of increase in numbers of developmental regulatory genes. Moreover, because even randomly generated networks commonly produced complex patterns, chance appearance of simple networks capable of pattern formation probably was not the rate limiting step for evolution of complex multicellular organisms.

Frequency and complexity of pattern formation depends on the distribution of network connections

How did the network connectivity affect the pattern formation? In networks, where on average 50% of the connections were negative, the pattern formation frequency was lower but the pattern complexity was higher than in networks where 50% of interactions were merely inactivated by setting their strength to 0 (Fig. 2B). This suggests that negative signaling interactions are essential for developmental increase in metazoan complexity. This also indicates that the

negative interactions do not act merely by passively replacing the positive connections, but by actively disrupting the existing positive circuits.

The overall effects of the negative or inactive interactions depended on their distribution within the networks. This is shown by the comparison of the 6-signal networks where on average 50% of the interactions were randomly inactivated (Fig. 2Ba,d), 50% of the interactions were negative (Fig. 2Bb,e), 50% of the interactions were eliminated and of the rest 50% were negative and 50% positive (Fig. 2Bc,f). In Fig. 2Ba,b,c, the signals and their effectors were differently regulated, whereas in d,e,f the regulatory interactions were identical for signals and their effectors. Interestingly, when the signals and their effectors were regulated by the same sets of positive, negative, or inactive connections, patterns were produced more frequently than when connections were distributed differently between signals and their effectors. Indeed, when signals and their effectors were similarly regulated, even at 50% connectivity, the pattern formation likelihood was higher than in fully connected networks where signals and their effectors were differently regulated (Fig. 2Bc,f). This suggests that similar regulation of whole signaling module (signal and its intracellular effector) reduces the disruptive effects of inactive or negative connections, whilst maintaining most of the complexity increasing effects of the negative interactions.

The complex patterns can be generated based on qualitative network structure only

Above results show that the same 4 - 12 genes can produce different patterns depending on the set of connections between the genes. Did the differences between the network outputs depend on the strengths of the individual interactions alone, or did the network structure have an effect on pattern formation? The role of network structure compared to the strengths of the individual interactions in creating complex patterns can be analyzed by controlling the strengths of the interactions while maintaining the same network structure. Decreasing the rate of gene decay allows longer time window for the gene product to act. Decreasing the interaction constants reduces the rate of product increase and/or decrease by the current signals. When I decreased the interaction constants by multiplying them all with 0.2, 0.4, 0.6 or 0.8, it reduced the maximum pattern size, complexity and pattern formation frequency (Fig. 3), as does reducing the amount of signal amplification (not shown). Decreasing the rate of decay by setting it to 0.4, 0.3, 0.2, 0.1 or 0.0 increased the maximum pattern size, complexity and pattern formation frequency (Fig. 3). Decreasing simultaneously both rate of decay and the interaction constants produced patterns at intermediate frequencies and pattern complexities, though the pattern sizes increased. This is in accordance with the earlier studies that indicate that the strengths of the connections (or kinetic parameters of the signaling interactions) affect the development of complex patterns (Asai et al., 1999; Wearing et al. 2000).

However, when the effective concentrations of the signals and/or their effectors were thresholded to give binary response, that changed the target gene concentration for the amount of the full interaction constant (if the concentration of the signal and/or the effector were above the threshold level) or for zero amount (if the concentration was below the threshold level), the networks could still often produce multiple cell types, although more often at lower than at higher

thresholds (Fig. 4A). Giving different thresholds for different signals and/or effectors produced complex patterns with lower frequency than giving the lowest threshold for all signals and/or effectors, but at higher frequency than giving the highest threshold for all signals and/or effectors. Moreover, complex patterns could also form, when all the negative interaction constants were set to -0.7 and all the positive interaction constants were set to 0.7 (Fig. 4Bs). Thresholding the standard interaction constant [-0.7, 0.7] networks also produced complex patterns (Fig. 4B). Together, the results indicate that the differences between the strengths of interactions are not necessary for producing complex patterns, nor are different concentration thresholds of one (or more) morphogen(s) required for different cell type responses in this model. The same twelve signaling genes could produce multiple gene combinations in different patterns even when only qualitative network structures were different. This strongly suggests that the signaling network structures must be under selection in evolution of multicellular lifeforms.

Non-periodic patterns

Many of the complex patterns (> 2 gene combinations) produced by these small, random networks were non-periodic or quasiperiodic, having both periodic and non-periodic elements (Fig. 5). This suggests that the evolution of positional co-ordinate systems for unique organs and cell types is easy. Though the induction of different cell types depended on cell-cell relays, many cell types were several cells wide in radial axis (e.g., Fig. 5). This recapitulates previous studieson binary fate choises by juxtacrine signaling (Wearing et al., 2000). Since some important intercellular signals do not diffuse over long distances (e.g., Collier et al., 1996; Juliano and Varner, 1993), it is possible that some long distance (non-periodic) patterning events creating positional information result from cell-to-cell information relays.

Incidentally, the random strength interactions, standard strength interactions, and random/standard strength interactions combined to binary signaling thresholds could produce also non- or quasiperiodic patterns (Fig. 5). Combined to the cell-cell relay that eliminated long distance diffusion from the model, this indicates that multiple concentration dependent responses to a diffusible morphogen are not in theory essential for development of a positional co-ordinate system (Wolpert, 1998; Gurdon and Bourillot, 2001). Rather, this relay model translates one initial local signal (a single morphogen peak) into a sequence of cells with different expression combinations via series of secondary inductions (Meinherdt and Gierer 1980). Because same genes and inputs were used for in different networks in a same simulation, the precense and complexity of a pattern depended on the signaling hierarchy between the genes.

Discussion

Evolution of structurally complex multicellular lifeforms, like metazoans, depends on heritable changes in morphogenesis. Morphogenesis is guided by a network of gene-gene interactions coded in the genome of the organism. However, relatively few studies have tackled the effects of network structure in developmental increase in complexity (e.g., Salazar-Ciudad et al., 2000, 2001b; Kauffman, 1993). Therefore, I wanted to know which basic properties of the

signaling networks affect the formation of complex expression patterns that guide the cell differentiation and organogenesis.

Complex patterns did not require complex genomes or complex input information

Metazoan complexity is often equated with their complex 13000 – 40000 gene genomes. However, the simulations showed that even simple 12-gene networks can commonly produce patterns with multiple different cell types from a single spatial input (Figs. 1 and 2). This conforms with the earlier studies, which showed that gene networks commonly produce single gene spatial patterns (Salazar-Ciudad et al. 2000). Thus, theoretically only small portion of the metazoan genome is required for organizing metazoan embryogenesis. Inversely, this supports the idea that the effects of many genes can be integrated to relatively few but mutually separable cellular responses, the combinations of which can produce an impressive range of morphological complexity (Gerhart and Kirschner, 1997; Salazar-Ciudad and Jernvall, 2002). Both inferences drastically simplify the problem of evolution of developmental complexity.

The key to the developmental complexity is the spatial and temporal interplay between individual cells that express different parts of the whole network. *In vivo*, the same cell may at any given moment have access to a different parts of its genome, depending on its current proteome that the cell uses for reading its genome. In these simulations, each iteration can produce a different output (Fig. 1). On the other hand, though almost all cells in the model are initially similar, being responsive to the same signal, after this initial inductive signal has passed, each cell's temporal differences relative to its neighbors (together with the continuing relay of new signals back and forth) can induce different sets of secondary responses. The cells *in vivo* may also be able to respond differently to multiple alternative initial input signals, and the temporal effects of signaling may be more complex. For example, because of the biochemical differences between signaling pathways, different signals may be produced, or they may be translated into various cellular responses at different rates (Gurdon and Bourillot, 2001; Tabata, 2001). However, despite the greater *in vivo* complexity of signaling, the effects of temporal differences between the target cells are likely to be analogous to those in this model. Hence, the simulations suggest that multiple spatially separate inputs are not, at least initially, essential for generating metazoan cell types.

Network structure, pattern formation, and evolution

Many of the previous studies have emphasized the effects of the strengths of the individual interactions on pattern formation (e.g., Asai et al., 1999; Wearing et al., 2000; Meir et al. 2002). However, both prokaryotic and eukaryotic gene regulatory networks characteristically contain many negative interactions (e.g., von Dassow and Odell, 2002; Gerhart and Kirschner, 1997, Davidson et al., 2002), and the scale-free distribution of connections between individual genes is often further non-randomly organized (Ravasz et al, 2002; Shen-Orr et al. 2002; Snel et al., 2002). These architectural features are likely to have been selected for their functionality.

Though the simulations showed that changing the strengths of the interactions also changed the pattern formation capabilities of the networks (Fig. 3), standardizing the interactions constants and thresholding the effective gene concentrations did not remove the networks ability to generate complex patterns (Fig. 4). This supports the view that the qualitative structure of the network has a role in generating developmental complexity (Salazar-Ciudad et al., 2000; Kauffman, 1993). Because individual signaling molecules and *cis*-regulatory enhancers or suppressors are, to an extent, interchangeable between species and even phyla despite some sequence divergence (Gerhart and Kirschner, 1997; Onuma et al., 2002; Ludwig et al., 1998), morphological evolution may have depended as much on qualitative changes in network architecture as on the quantitative changes in the strengths of individual interactions.

The distribution of connections dramatically affects the pattern formation frequencies and the development of complex patterns. Though the pattern formation capabilities of the network depended on its size, presence of negative interactions was more important (Fig. 2). Random negative connections increase the pattern complexity but decrease the pattern formation frequencies. However, without negative interactions only simple patterns were produced (Fig. 2), especially in the absence of expression decay (not shown). Thus, the appearance direct negative interactions, e.g., DNA-binding proteins that block the promoter activity, or indirect negative interactions that downregulate the upstream amplification of signals leading to positive interactions, has been crucial for evolution of complexity.

Inactive connections decrease the pattern formation frequencies and the pattern complexity. This happens also in networks that contain negative interactions (Fig. 2). However, when both extracellular signals and their intracellular effectors were similarly regulated by same signals, the pattern formation likelihoods increased both in fully connected and in 50% connected networks. The distribution of connections in both metabolic and transcriptional networks is known to be more or less scale-free (Wolf et al., 2002, Goh et al. 2002) and non-random or modular (Ravasz et al., 2002; Shen-Orr et al., 2002; Snel et al. 2002; Guelzim et al., 2002), probably because of the selection acting on the network structure. Because maintaining the expression of regulatory genes is often essential for developmental pattern formation, and similar regulation of both signal and its effector(s) within a regulatory signaling module increases the pattern formation likelihood of the network (Fig. 2B), also autoregulatory loops and/or co-regulated cassettes of developmental regulatory genes are likely to have been selected for during evolution. In prokaryote and yeast metabolic networks, co-regulation of metabolic pathway components is frequent by common operons and/or common transcription factors (Struhl, 1999; Ihmels et al., 2002). Therefore, prevalence of co-regulation within analogous metazoan developmental signaling networks would not be surprise. If, however, further analysis of *in vivo* signaling network structures shows that co-regulation of metazoan signaling pathway components occurs only at low frequency, this may indicate opposing evolutionary forces.

Quasiperiodic and non-periodic patterns

Though self-organizing patterns that produce periodically self-similar outcomes are more commonly studied (Asai et al., 1999; von Dassow et al., 2000), real animal bodyplans are much more complex, combining many non-periodic to

many periodic features (Gerhart and Kirschner, 1997). The non-periodic pattern elements can be biologically meaningful in positioning unique cell types and organs (Kauffman, 1993). Non-periodic pattern formation is also essential for morphological diversification of serially repeated structures, such as plant organs (Kieffer and Davies, 2001), metazoan segments (McGinnis and Kuziora, 1994; Gilbert, 1998), or vertebrate teeth (Stock et al., 1997). Because all the different pattern elements must be produced by the same genome, the main questions are; 1) how complex a signaling network is required, and 2) can all patterns be produced by the same network, or does the genome contain multiple, parallel signaling networks?

The complex patterns produced in these simulations were common, and often non-periodic, or quasiperiodic, containing both periodic and non-periodic elements (Fig. 5). This demonstrates that simple genomes can produce all kinds of patterns.

The existence of non-periodic patterns shows that determination of multiple unique positional identities along the bodyaxes is possible even with small signaling networks, assuming that different combinations of the same genes regulate different cell fates or morphogenetic effects. Because the model was based on cell-to-cell relays and because the binary thresholds combined to standard strength interactions still produced quasi- or non-periodic patterns (Fig. 5), morphogen gradients with multiple different effective response thresholds (Wolpert, 1998; Gurdon and Bourillot, 2001) are not theoretically essential for non-periodic developmental pattern formation. In fact, because non-periodic and quasiperiodic patterns were common in these simulations, and largely dependent on intercellular signaling network structure, if morphogen gradients with differential response thresholds are a dominant mechanism for producing positional co-ordinates, they probably have some advantage, like pattern stability (Hunding et al. 1990), that has been favored in evolution of developmental signaling networks (Salazar-Ciudad et al. 2001b).

The existence of quasiperiodic patterns (Fig. 5) suggests that segregation into periodic and non-periodic pattern types is neither clear nor, at least in theory, necessary. This means that to better integrate the different types of patterns (periodic, quasiperiodic, and non-periodic ones) into the same body plan, further studies signaling network function and evolution are required. For example, because random chance in these simulations commonly produced signaling networks that generate quasiperiodic patterns, if the periodic and the non-periodic pattern elements *in vivo* are predominantly produced by different developmental modules, this feature in signaling network structure is likely to have been selected for during evolution.

Conclusions

Real animal bodyplans require nonperiodic patterns, e.g., for positional co-ordinate systems (Wolpert, 1998; Kauffman, 1993). The simulations show that even small signaling networks can commonly generate multiple cell types from very simple inputs. Because (unlike large signaling networks) small signaling networks are reasonably likely to arise by chance alone, the evolution of metazoan complexity and positional co-ordinate systems was not limited by network

complexity. This suggests that the evolution of metazoan complexity has not depended on the genome size alone (see also Claverie, 2001; Salazar-Ciudad et al, 2001b; Gerhart and Kirschner, 1997).

Though the strengths of interactions have a role in pattern formation, differences in interactions strengths or multiple concentration dependent response thresholds to a diffusible morphogen are not essential for generating spatial patterns. Thus, the same 4 – 12 genes could produce very different patterns depending on the structure of the signaling network alone. In other words, the qualitative structure of the signaling network alone has a significant input on the development of complexity (Salazar-Ciudad et al., 2000).

The patterns on these simulations are different subsets of same larger set of genes. However, with combinatorial control logic, different subsets of same relatively small set of developmental regulatory genes can be translated into multiple organs or cell types. Interestingly, the simulations show that many of the gene combination patterns based on cell-to-cell relays were spatially nonperiodic or quasiperiodic with overlapping periodic and nonperiodic features. Hence, multiple networks are not theoretically essential for generating spatial co-ordinate systems in complex multicellular organisms.

Negative interactions greatly increased the pattern complexity. The frequency and the complexity of the patterns increased if only the effectors were negatively regulated or if both signals and their effectors are similarly regulated. Thus, there may be an evolutionary bias for co-regulated modules in developmental signaling networks. There also may be a bias for controlling the complexity via the intracellular signal processing rather than by regulating the pattern of extracellular signals. Because the patterning complexity depends on the structure of the signaling network, and because the network has to evolve by random mutations, it may be that the evolution of multicellularity has also been facilitated by evolution of intracellular requlatory mechanisms that predispose the mutations for negative interactions and co-regulation.

To test these hypotheses, more studies are needed on the effects of random mutations on networks that use different kinds of signal processing logic, and on the distributions of connections in the real metazoan signaling networks.

Methods

1000 networks of each kind were generated and tested with same initial conditions. The 4-, 6-, 8-, 10- or 12-gene networks consisted of the same 2-, 3-, 4-, 5- or 6-signals and their specific intracellular effectors, through which they only could interact. However, the 8-, 12-, 16-, 25- or 36-interaction constants between each signal and all its target genes for each network were randomly set between [-1, 1] before simulations. On average 50% of the connections were negative (repressive) and 50% were positive (inductive) Each gene could be expressed in each cell at strength [0, 4]. The signals deterministically increased or decreased the concentrations of their targets at each iteration, depending on

the current concentration of the received signal multiplied with the current concentration of the effector in the receiving cell, and the constant interaction strength.

The simulations were run in artificial organisms made from matrix of 30 x 40 cells. Each cell signals to itself at strength 1.0 x signal concentration, and communicates with four neighbors at strength 0.25 x signal concentration each, leading to cell-to-cell relay of signals. Each signal is amplified by the concentration of its effector in each target cell. The opposing edges were joined creating a continuous 2D surface. The pattern formation began from an initial input of one signal present at full concentration in one cell. Additionally, all cells initially expressed the receptor for the original signal at full concentration. The expression of all genes decayed at specific rate 0.5 (unless different rate was specified) per iteration. The system was allowed to run 50 or 150 iterations before analyzing the output.

I tested the effects of connectivity distribution with 6-signal networks. The average connectivity was reduced by replacing interactions with 0-strength interactions. Both fully random distributions of 0-strength and/or negative connections and random controls that were identical for signal and its effector were tested. If only signals were also negatively controlled, all interactions controlling the effector were positive, and vice versa. In standard interaction experiments, all the positive interactions were set to 0.7 and negative interactions at -0.7. In induction strength experiments, the strengths of the interactions were reduced by multiplying them with 0.2 – 1.0. In threshold experiments, above the threshold concentrations, the signals and/or their effectors affected their targets for the (non-multiplied) amount of interaction strengths coded in the network, whilst below threshold concentrations they had no effect.

A pattern was defined as an virtual organism where different cells expressed different on/off-combinations of genes. Some patterns contained only one gene combination in some cell(s) and cells without expression. The patterns were analyzed by counting the number of networks out of 1000, that after the final iteration had produced spatial patterns, the sizes of the produced patterns, and the numbers of cell types.

The simulation and analysis programs were written in PERL 5.6 and used GD-, PDL-, and PDL::IO::FastRaw-modules (http://www.cpan.org). The simulations were run in MacOSX G4 PowerMac (Apple). The column plots were made with Excel (Microsoft). The image plates were made in Photoshop (Adobe) and Powerpoint (Microsoft). The PERL-scripts are available at request from author.

Acknowledgements

Thanks to M. Ronshaugen, B.T. Holland, B. Inwood, stimulating discussions, B. Inwood, I. Ovcharenko for help with computers, M. Biggin, B.T. Holland for critical reading, W. McGinnis, M.D. Biggin, Academy of Finland, Helsingin Sanomain 100-vuotissäätiö, NIH for funding. This work was performed at the Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC03-76SF00098, University of California.

References

ASAI, R., TAGUCHI, E., KUME, Y., SAITO, M., KONDO, S. (1999). Zebrafish *Leopard* gene as a component of the putative reaction-diffusion system. *Mech. Dev.* **89 (1-2),** 87-92, doi:10.1016/S0925-4773(99)00211-7.

BACHVAROVA, R. F. (1999). Establishment of anterior-posterior polarity in avian embryos. *Curr. Opin. Genet. Dev.* **9 (4),** 411-416, doi:10.1016/S0959-437X(99)80062-8.

BORNHOLDT, S. (2001). Modeling genetic networks and their evolution: a complex dynamical systems perspective. *Biol. Chem*. **382 (9),** 1289-1299.

BOSENBERG, M.W., MASSAGUE, J. (1993). Juxtacrine cell signaling molecules. *Curr. Opin. Cell Biol.* **5 (5),** 832-838.

CLAVERIE, J. M. (2001). Gene number. What if there are only 30,000 human genes? *Science* **291 (5507),** 1255-1257.

COLLIER, J. R., MONK, N. A. M., MAINI, P.K., LEWIS, J. H. (1996). Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *J. Theor. Biol*. **183 (4),** 429-446, doi:10.1006/jtbi.1996.0233.

VON DASSOW, G., ODELL, G. M. (2002). Design and constraints of the Drosophila segment polarity module: robust spatial patterning emerges from intertwined cell state switches. *J. Exp. Zool*. **294 (3),** 179-215.

DAVIDSON, E. H, RAST, J. P., OLIVERI, P., RANSICK, A., CALESTANI, C., YUH, C. H., MINOKAWA, T., AMORE, G., HINMAN, V., ARENAS-MENA, C., ET AL. (2002). A genomic regulatory network for development. *Science* **295 (5560),** 1669-1678.

ELOWITZ, M. B., LEVINE, A. J., SIGGIA, E. D., SWAIN, P. S. (2002). Stochastic gene expression in a single cell. *Science* **297 (5584),** 1183-1186.

GERHART, J., KIRSCHNER, M. (1997). *Cells, embryos, and evolution*. Malden, Mass: Blackwell Science.

GILBERT, S. F. (1997). *Developmental biology, 5th ed.* Sunderland, Mass: Sinauer Associates.

GOH, K.-I., OH, E., JEONG, H., KAHNG, B., KIM, D. (2002). Classification of scale-free networks. *Proc. Natl. Acad. Sci. USA* **99 (20),** 12583-12588.

GUELZIM, N., BOTTANI, S., BOURGINE, P., KEPES, F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* **31 (1),** 60-63.

GURDON, J. B., BOURILLOT, P.-Y. (2001). Morphogen gradient interpretation. *Nature* **413 (6858),** 797-803.

HUNDING, A. KAUFFMAN, S. A., GOODWIN, B. C. (1990). *Drosophila* segmentation: Supercomputer Simulation of

Prepattern Hierarchy. *J. Theor. Biol.* **145 (3),** 369-384.

HUNTER, T. (2000). Signaling – 2000 and Beyond. *Cell* **100 (1),** 113-127.

IHMELS, J., FRIEDLANDER, G., BERGMANN, S., SARIG, O., ZIV, Y., BARKAI, N. (2002). Revealing modular organization in the yeast transcriptional network. *Nature Gen.* **31 (4),** 370-377.

JULIANO, R. L., VARNER, J. A. (1993). Adhesion moleculas in cancer: the role of integrins. *Curr. Opin. Cell Biol*. **5 (5),** 812-818.

KAUFFMAN, S. A. (1993). *The Origins of Order.* New York: Oxford University Press.

KIEFFER, M., DAVIES, B. (2001). Developmental programmes in floral organ formation. *Semin. Cell Dev. Biol.* **12 (5),** 373-380. doi:10.1006/scdb.2001.0266.

LUDWIG, M. Z., PATEL, N. H., KREITMAN, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. *Development* **125 (5),** 949-958.

MCGINNIS, W., KUZIORA, M. (1994). The Molecular Architects of Bodyplan. *Sci. Am*. **270 (2),** 58-61, 64-66.

MEINHARDT, H., GIERER, A. (1980). Generation and Regeneration of Sequence of Structures During Morphogenesis. *J. Theor. Biol.* **85 (3),** 429-450.

ONUMA, Y., TAKAHASHI, S., ASASHIMA, M., KURATA, S., GEHRING, W. J. (2002). Conservation of Pax 6 function and upstream activation by Notch signaling in eye development of frogs and flies. *Proc. Natl. Acad. Sci. USA* **99 (4),** 2020-2025.

OSTER, G. F., MURRAY, J. D. (1989). Pattern Formation Models and Developmental Constraints. *J. Exp. Zool.* **251 (2),** 186-202.

RAVASZ, E., SOMERA, A. L., MONGRU, D. A., OLTVAI, Z. N., BARABASI, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* **297 (5586),** 1551-1555.

RUBIN, G. M., YANDELL, M. D., WORTMAN, J. R., GABOR MIKLOS, G. L., NELSON, C. R., HARIHARAN, I. K., FORTINI, M. E., LI, P. W., APWEILER, R., FLEISCHMANN, W., ET AL. (2000). Comparative genomics of the eukaryotes. *Science* **287 (5461),** 2204-2215.

SALAZAR-CIUDAD, I., GARCIA-FERNANDEZ, J., SOLÉ, R. V. (2000). Gene networks capable of pattern formation: from induction to reaction-diffusion. *J. Theor. Biol.* **205 (4),** 587-603, doi:10.1006/jtbi.2000.2092.

SALAZAR-CIUDAD, I., SOLÉ, R. V., NEWMAN, S. A. (2001a). Phenotypic and dynamical transitions in model genetic networks. II. Application to the evolution of segmentation mechanisms. *Evol. Dev.* **3 (2),** 95-103.

SALAZAR-CIUDAD, I., NEWMAN, S. A., SOLÉ, R. V. (2001b). Phenotypic and dynamical transitions in model genetic

networks. I. Emergence of patterns and genotype-phenotype relationships. *Evol. Dev.* **3 (2),** 84-94.

SALAZAR-CIUDAD, I., JERNVALL, J. (2002). A gene network model accounting for development and evolution of mammalian teeth. *Proc. Natl. Acad. Sci. USA* **99 (12),** 8116-8120.

SHEN-ORR, S. S., MILO, R., MANGAN, S., ALON, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Gen.* **31 (1),** 64-68

SNEL, B., BORK, P., HUYNEN, M. A. (2002). The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci. USA* 99 (9), 5890-5895.

SOBOTT, F., ROBINSON, C. V. (2002). Protein complexes gain momentum. *Curr. Opin. Struct. Biol.* **12 (6),** 729-734, doi:10.1016/S0959-440X(02)00400-1.

STOCK, D. W., WEISS, K. M. ZHAO, Z. (1997). Patterning of the mammalian dentition in development and evolution. *BioEssays* **19 (6),** 481-490.

STRUHL, K. (1999). Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. *Cell* **98 (1),** 1-4.

TABATA, T. (2001). Genetics of morphogen gradients. *Nature Rev. Gen.* **2 (8),** 620-630.

TURING, A. M. (1952) *Philos. trans. R. Soc. Ser. B* **237 (641),** 37-72

WEARING, H. J., OWEN, M. R., SHERRAT, J. A. (2000). Mathematical modelling of juxtacrine patterning. *Bull. Math. Biol.* **62 (2),** 293-320, doi:10.1006/bulm.1999.0152.

WOLF, Y. I., KAREV, G., KOONIN, E. V. (2002). Scale-free networks in biology: new insights into the fundamentals of evolution? *BioEssays* **24 (2),** 105-109.

WOLPERT, L. (1998). *Principles of Development.* New York: Oxford University Press.

Figure legends

Figure 1.

First 15 iterations from initial input of one signal in a fully connected random 6-signal network with negative connections at decay 0.5.

Figure 2.

A) Effects of network size and negative interactions. Cumulative numbers of 2-, 3-, 4-, 5-, and 6-signal networks expressing patterns of different complexities after 50 iterations at decay 0.5. *n100 fully connected with 50% connections negative; p50 fully 50% connected, positive connections only.*

B) Effects of distribution of positive and negative interactions. *a)* 50% connected networks with fully random positive interactions, *b)* fully connected networks with fully random positive and negative interactions, *c)* 50% connected networks with fully random positive and negative interactions, *d)* 50% connected networks with similar positive interactions for both signals and their effectors, *e)* fully connected networks with similar positive and negative interactions for signals and their effectors, f) 50% connected networks with similar positive and negative interactions for signals and effectors. The decay was 0.5, number of iterations 50.

Figure 3.

Effects of product decay and interaction constants on output of fully connected 6-signal networks. Rates of decay were 0.5 (*a, g – l*), 0.4 (*b, m*), 0.3 (*c, n*), 0.2 (*d, o*), 0.1 (*e, p*), or 0.0 (*f*) per iteration. The interaction constants were multiplied with 1.0 (*a – g, l*), 0.8 (*h, m*), 0.6 (*i, n*), 0.4 (*j, o*), or 0.2 (*k, p*) to make the effects partially comparable to decay rates 0.4, 0.3, 0.2, 0.1. The patterns were allowed to develop for 50 iterations

Figure 4.

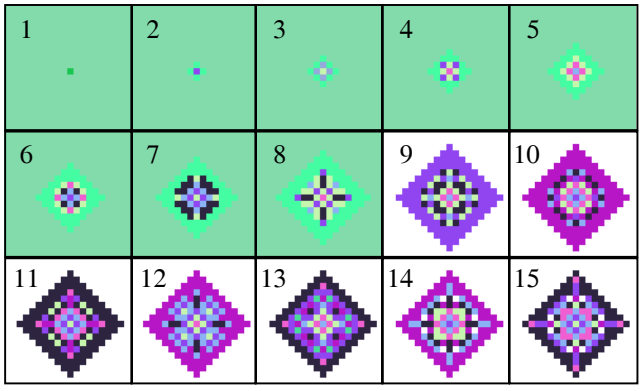
A) Thresholding the effective concentrations of all the signals (*a, b, c, d, e, f*), effectors (*g, h, i, j, k, l*) or signals * receptors (*m, n, o, p, q, r*) to 0.2 (*e, k, q*), 0.4 (*d, j, p*), 0.6 (*c, i, o*), 0.8 (*b, h, n*) or 0.1 (*a, g, m*) reduced the pattern formation frequency and complexity relative to non-thresholded simulations (*s*) as the threshold size increased. Different thresholds for different signals (*f*), effectors (*l*), or signals * receptors (*r*) produced patterns at frequency and complexity intermediate to the lowest and highest thresholds.

B) Standardizing the interaction constants to [-0.7, 0.7] increased the pattern formation frequency relative to random networks outputs. The outcomes of different conditions were also more equal to each other. The rate of decay in both simulations was 0.5, the number of iterations 50. *a – s; as above*

Figure 5.

Examples of patterns produced by 6-signal networks with random strength interactions (A), standard strength interactions (B), random interactions at threshold concentration 0.2 for signal * effector complex (C), and standard strength interactions at same threshold conditions (D). The patterns developed for 150 iterations at decay 0.2.

Figure 1



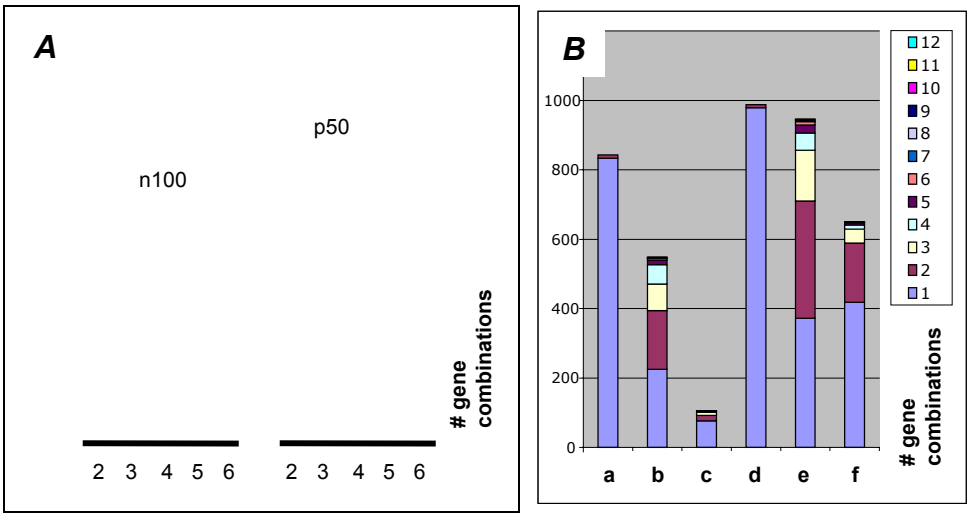

A
p50
n100
2 3 4 5 6
2 3 4 5 6
gene combinations
B
1000
800
600
400
200
0
a b c d e f
12
11
10
9
8
7
6
5
4
3
2
1
gene combinations

Figure 2.

Figure 3.

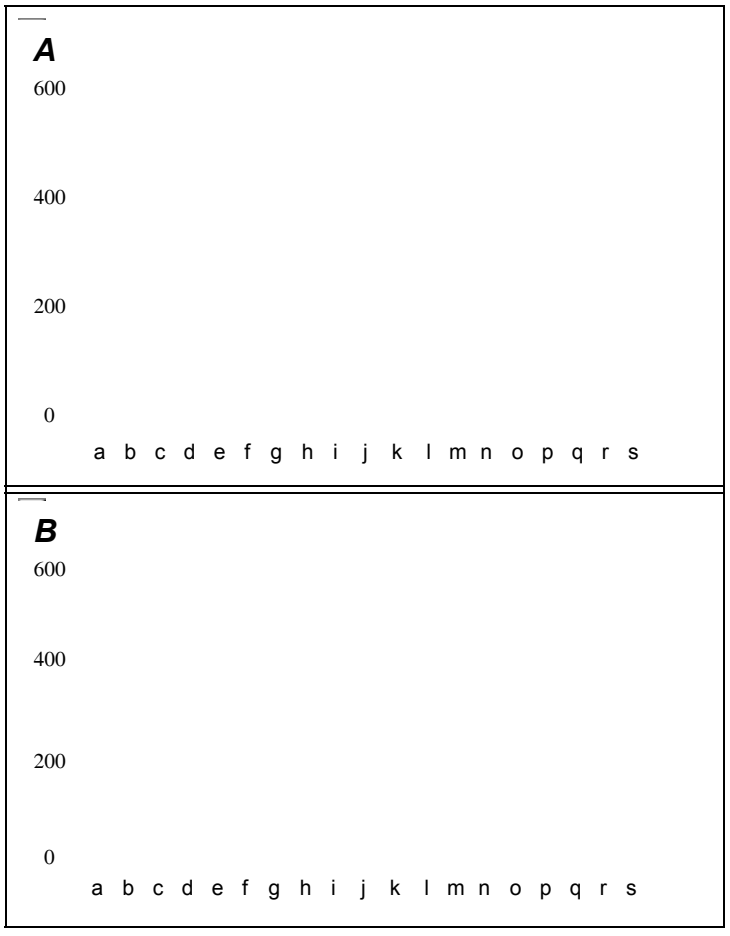

A
600
400
200
0
a b c d e f g h i j k l m n o p q r s
B
600
400
200
0
a b c d e f g h i j k l m n o p q r s

Figure 4.

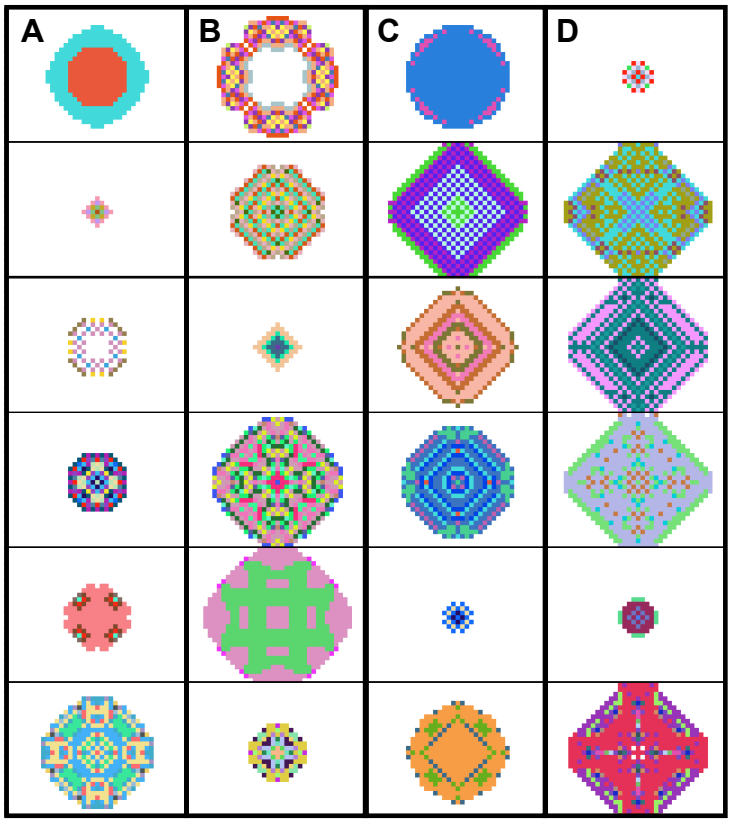


Figure 5.